

# Attentional Capacity Limits Gap Detection during Concurrent Sound Segregation

Ada W. S. Leung<sup>1,2</sup>, Pierre Jolicoeur<sup>3,4,5,6</sup>, and Claude Alain<sup>2,7</sup>

## Abstract

■ Detecting a brief silent interval (i.e., a gap) is more difficult when listeners perceive two concurrent sounds rather than one in a sound containing a mistuned harmonic in otherwise in-tune harmonics. This impairment in gap detection may reflect the interaction of low-level encoding or the division of attention between two sound objects, both of which could interfere with signal detection. To distinguish between these two alternatives, we compared ERPs during active and passive listening with complex harmonic tones that could include a gap, a mistuned harmonic, both features, or neither. During active listening, participants indicated whether they heard a gap irrespective of mistuning. During passive listening, participants watched a

subtitled muted movie of their choice while the same sounds were presented. Gap detection was impaired when the complex sounds included a mistuned harmonic that popped out as a separate object. The ERP analysis revealed an early gap-related activity that was little affected by mistuning during the active or passive listening condition. However, during active listening, there was a marked decrease in the late positive wave that was thought to index attention and response-related processes. These results suggest that the limitation in detecting the gap is related to attentional processing, possibly divided attention induced by the concurrent sound objects, rather than deficits in preattentive sensory encoding. ■

## INTRODUCTION

The auditory environment is often composed of a myriad of sound sources, all of which compete for our attention. According to the object-based account of auditory attention (Shinn-Cunningham, 2008; Alain & Arnott, 2000), a predominantly automatic segregation process results in the formation of objects that are subsequently stored in STM. The notion of “auditory object” refers to a grouping of sounds into a coherent whole or stream such that it seems to emanate from a single acoustic source (Alain, 2007). According to the auditory scene analysis account (Bregman, 1990), incoming concurrent sounds are initially analyzed and then perceptually grouped into distinct sound objects based on the physical properties of the sound, such as common onsets and offsets, harmonic structure, and continuity of frequency over time (Bregman, 1990). These sound objects form the basic units for attentional selection (Dyson & Ishfag, 2008; Shinn-Cunningham, 2008; Alain & Arnott, 2000). Evidence from behavioral and EEG studies demonstrated that sequential (Snyder & Alain,

2007; Cusack, Carlyon, & Robertson, 2000) and concurrent (Dyson, Alain, & He, 2005; Alain & Izenberg, 2003) sound segregation can occur irrespective of a listener’s attention. Once the auditory scene has been partitioned into distinct sound objects, a selection process allows an individual to focus or switch their attention from one sound object to another (Backer & Alain, 2014). Although there is good evidence that the perception of sound stimuli is influenced by factors like stimulus salience or behavioral goals and that these factors likely interact to determine which sound(s) are perceived, attended, and remembered, how these processes interact when multiple sound objects compete for attention remains unknown.

Recently, we showed that the perception of a gap (i.e., a brief silence) inserted into a complex harmonic sound was more difficult when it included a mistuned component that “popped out” as a separate auditory object (Leung, Jolicoeur, Vachon, & Alain, 2011). This effect was observed even for gap durations well above the usual perceptual threshold (for sounds that do not contain a mistuned component). Leung et al. proposed that perceiving the gap was more difficult because the attention was divided between two competing sound objects, thereby interfering with the detection of the gap. However, other reasons could explain why detecting the gap was more difficult during the presence of an inharmonic component (Heinrich, Alain, & Schneider, 2004). One could imagine that the presence of a mistuned harmonic in an otherwise in-tune harmonic complex could reduce

<sup>1</sup>University of Alberta, <sup>2</sup>Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Canada, <sup>3</sup>Université de Montréal, <sup>4</sup>Centre de Recherche en Neuropsychologie et Cognition (CERNEC), Montréal, Canada, <sup>5</sup>BRAMS (International Laboratory for Brain, Music, and Sound Research), Montréal, Canada, <sup>6</sup>Centre de Recherche de l’Institut Universitaire de Gériatrie de Montréal (CRIUGM), <sup>7</sup>University of Toronto

gap detection because of low-level peripheral interactions such as beating, auditory induction, perceptual illusion, or restoration, which are thought to result from the reorganization of acoustic energy across multiple wave bands and phases (Recanzone & Sutter, 2008). From behavioral data alone, it is difficult to determine whether the gap detection impairment is because of impoverished sensory or perceptual processing of the sound stimuli or a division of attention. That is, deficits in gap detection when two sound objects are simultaneously presented may be because of a failure in encoding or insufficient processing at a higher cognitive level.

The recording of scalp ERPs provides a mean to assess the automatic encoding of sound features such as harmonicity and gap as well as higher cognitive processes such as attention and memory. Prior research has revealed neural correlates for both concurrent sound perception and gap detection. For instance, the perception of concurrent sound objects is associated with an enhanced negativity that overlaps with the N1 and P2 waves elicited by sound onset (Dyson & Alain, 2004; Alain, Schuler, & McDonald, 2002; Alain, Arnott, & Picton, 2001). This enhanced negativity, referred to as the object-related negativity (ORN), is most prominent over the frontocentral scalp region and is best isolated by a difference wave between the ERPs elicited by tuned and mistuned stimuli (such as those including a component mistuned by 16% that clearly pops out of the complex; Alain, 2007; Alain et al., 2001). With respect to gap detection, short duration sounds (e.g., 200 msec) that include even shorter gaps in the middle generate smaller sensory evoked responses (i.e., N1 and/or P2 waves) relative to stimuli without a gap (Ross, Schneider, Snyder, & Alain, 2010; Heinrich et al., 2004; Hillyard & Picton, 1978). Gap-related neural activity can be isolated by subtracting auditory ERPs elicited by sounds with and without the gap. Both the ORN and the gap-related activity can even be observed when listeners are performing another task (e.g., reading a book, watching a movie; Alain, 2007; Alain et al., 2001).

Interestingly, when participants are actively processing a mistuned harmonic or gap, the perceptual decision regarding the stimuli elicits a late positive complex (LPC) that peaks at about 600 msec over the parietal scalp region in addition to the ORN or gap-related activity. This component has been associated with conscious identification of an auditory event and is indicative of attentional processes (Alain et al., 2002; Martin, Sigal, Kurtzberg, & Stapells, 1997; Parasuraman, Richer, & Beatty, 1982). The amplitude of the LPC is strongly modulated by attention and target detectability (e.g., Dell'Acqua et al., 2015; Picton, 1992). Many studies have reported that the mean amplitude and mean area of the LPC are associated with behavioral responses in attention and STM tasks (Xu, Zhang, Ouyang, & Hong, 2013; Wolk et al., 2006; Curran, Schacter, Johnson, & Spinks, 2001).

In this study, we measured auditory ERPs during both passive and active listening conditions to examine whether

harmonicity and gap-related activities interact during the encoding and attentional processes. By examining the N1 and P2 waves as well as the attention-dependent LPC, one could determine whether gap detection impairments in the previous study (Leung et al., 2011) were because of failure of low-level sensory encoding or the division of attention between concurrent sound objects. If the deficits in gap detection were because of limitations during the encoding process, then the amplitude of sensory evoked responses elicited by the gap stimuli should be modulated by mistuning. Conversely, if the deficits in gap detection were because of a division of attention, then one would expect greater LPC amplitudes for gap detection in tuned compared with mistuned stimuli as well as correct compared with incorrect responses. Furthermore, if the ORN indexes the perception of concurrent sound objects, then one would expect greater ORN amplitudes for incorrect than correct responses (i.e., deficits in hearing the gap imply greater likelihood of hearing the mistuned harmonic as a separate object).

## METHODS

### Participants

Twenty right-handed young adults (mean age =  $24.30 \pm 4.27$  years, 10 women) gave informed consent and participated in the study. All participants had pure tone thresholds below 30 dB HL for frequencies ranging from 250 to 8000 Hz. None of them had neurological or psychological illnesses or were taking medication at the time of the experiment. The study was approved by the research ethics board of the Toronto Academic Health Sciences Network and the University of Toronto Human Subject Review Committee. Participants received \$25 in compensation for their participation in the study.

### Stimuli and Task

Stimuli consisted of four different complex sounds (200 msec in duration, 2.5-msec rise and fall time) generated by adding 10 pure tones of equal intensity. All stimuli had a fundamental frequency ( $f_0$ ) of 200 Hz. For half of the stimuli, which were referred to as "tuned," all tonal elements were an exact integer multiple of  $f_0$  (i.e., 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, and 2000 Hz). For the other half of the stimuli, which were referred to as "mistuned," the fourth tonal element ( $4 \times f_0$ ) was shifted upward by 16% of the original frequency (i.e., 928 Hz instead of 800 Hz). For each of the tuned and mistuned harmonics, half of the stimuli included a 15-msec gap that had a 2.5-msec rise and fall time and a 10-msec zero-amplitude plateau inserted in the middle of the third tonal element (600 Hz). The sounds were generated digitally at a sampling rate of 48.8 kHz using the System 3 Real-Time Processor from Tucker Davis Technologies RP-2 real-time processor (Alachua, FL). They were presented

binaurally through insert earphones (ER-3A; Etymotic Research, Elk Grove, CA) at 68-dB sound pressure level as measured with a sound pressure level meter using an artificial ear (Dalimar Instrument Inc., Quebec, Canada).

All participants took part in a passive listening condition and an active listening condition. A total of 1200 stimuli were presented in each condition. The passive condition consisted of two blocks of 600 stimuli presented in random order (150 stimulus presentations  $\times$  4 stimulus types). The ISI varied randomly between 1800 and 2200 msec (100-msec steps, rectangular distribution). The active condition consisted of six blocks of 200 stimuli presented in random order (50 stimulus presentations  $\times$  four stimulus types). The ISI varied between 2300 and 2700 msec (100-msec steps, rectangular distribution) to accommodate the participant's response on each trial. We used a custom Matlab program with Psychophysical Toolbox (version 11.0; The MathWorks, Natick, MA) on a Dell Pentium 4 PC with a SoundBlaster Live sound card (Creative Technology, Ltd.) to run the experiment.

All participants completed the passive listening condition before completing the active listening condition. During passive listening, participants watched a silent subtitled movie of their choice while the auditory stimuli were presented. Before participants started the active listening condition, they were given a chance to familiarize themselves with the stimuli by listening to each stimulus played six times consecutively in the following order: tuned, tuned gap, mistuned, and mistuned gap. During active listening, the participants indicated the presence or absence of the gap for each stimulus by pressing 1 or 3 on a computer keypad, respectively. Participants were given rest breaks between blocks. The experiment took place in a sound-attenuating chamber.

### Recording of Neuroelectric Brain Activity

The EEG was digitized continuously (sampling rate = 500 Hz) from an array of 64 electrodes with a bandpass filter of 0.05–100 Hz using NeuroScan Synamps2 (Compumedics, El Paso, TX). Eye movements were monitored with electrodes placed at the outer canthi and at the inferior orbits. During recording, all electrodes were referenced to the midline central electrode (i.e., Cz). For offline data analysis, they were rereferenced to an average reference. The analysis epoch consisted of 200 msec of prestimulus activity and 800 msec of poststimulus activity. For each participant, a set of ocular movements was obtained before and after the experiment (Picton et al., 2000). From this set, averaged eye movements were calculated for both lateral and vertical eye movements as well as for eye blinks. A PCA of these averaged recordings provided a set of components that best explained the eye movements. The scalp projections of these components were then removed from the experimental ERPs to minimize ocular contamination, using

BESA 5.2.0. Epochs contaminated by excessive deflections (greater than  $\pm 100 \mu\text{V}$  anywhere in the epoch) after correcting for ocular contaminations were excluded from the averages. For each participant, the remaining epochs were averaged according to electrode position, stimulus type (i.e., tuned, no gap; tuned, with gap; mistuned, no gap; and mistuned, with gap), experimental condition (i.e., active and passive listening conditions), and gap detection performance (i.e., correct rejection, hit, false alarm, and miss responses) using BESA. The ERPs were digitally filtered to attenuate frequencies above 30 Hz (12 dB/Oct, zero phase).

### Data Analysis

#### Behavioral Data

A sensitivity index ( $d'$ ) and response bias ( $\beta$ ) were calculated from hit and false alarm rates of gap detection in each of the two conditions (tuned and mistuned harmonics) during the active listening condition (Macmillan & Creelman, 1991). Paired  $t$  tests were performed on the  $d'$  and  $\beta$  values as well as the hit and false alarm rates, and RT, to compare gap detection performance as a function of harmonicity (tuned vs. mistuned).

*Analyzing sensory encoding from 0 to 300 msec.* The analysis of sensory encoding focused on comparing the magnitude of the evoked potential for the mistuning and gap-related component. A four-way repeated-measures ANOVA was computed and included the following within-subject factors: condition (active vs. passive), harmonicity (tuned vs. mistuned), gap (presence vs. absence), and site (i.e., frontocentral electrodes: FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, and CP2). These nine electrodes were selected because they best captured the evoked responses corresponding to the mistuning and the gap-related components (Heinrich et al., 2004; Alain et al., 2001). The ERPs for the mistuning and gap conditions were quantified as the mean amplitude between 100 and 200 msec and between 175 and 275 msec, respectively. These time windows were chosen based on prior research using comparable stimuli. The effect of mistuning typically peaks at about 150 msec after sound onset (Alain et al., 2001, 2002). The gap-related activity peaks at about 125 msec after gap onset (Campbell & Macdonald, 2011; Heinrich et al., 2004), which would correspond to about 225 msec after sound onset as the gap onset was at about 100 msec in this study. This 100-msec adjustment is not applicable to the mistuning component as the onset of the mistuning is the same as the onset of the sound.

*Analyzing attention during gap detection from 0 to 300 msec.* The first analysis was performed to examine the perception of the mistuned harmonic in the complex sound during the active listening condition by extracting the ORN from the ERP waveforms. This analysis allows us

to determine whether mistuning has an impact on the detection of a gap in the sound stimulus, that is, we would expect a significantly larger ORN for miss responses if the impairment of detecting the gap is because of the interference of the mistuning. The ORN was obtained by computing a difference wave between the tuned and mistuned stimuli. Mean ORN amplitude was extracted between 100 and 200 msec (Alain et al., 2001, 2002). For the ORN analysis, a three-way repeated-measures ANOVA was computed, which included the within-subject factors Gap, Response type, and Site.

A second analysis was performed to determine whether there was an interaction between gap detection accuracy and harmonicity. A three-way repeated-measures ANOVA on gap-related activity (mean amplitude between 175 and 275 msec) was computed, which included the factors Gap detection response (i.e., hit vs. miss), Harmonicity (i.e., tuned vs. mistuned), and Site.

*Analyzing attention during gap detection from 400 to 800 msec.* The LPC was quantified using the total positive area between 400 and 800 msec poststimulus onset. The mean area-under-the-curve provides a better estimate of the LPC than peak amplitude at the individual level (Luck, 2014), and the use of this measure, or the mean amplitude in a given window, is common (e.g., Dell'Acqua et al., 2015; Hasko, Groth, Bruder, Bartling, & Schulte-Korne, 2013). A four-way repeated-measures ANOVA was calculated using the factors Harmonicity (tuned vs. mis-

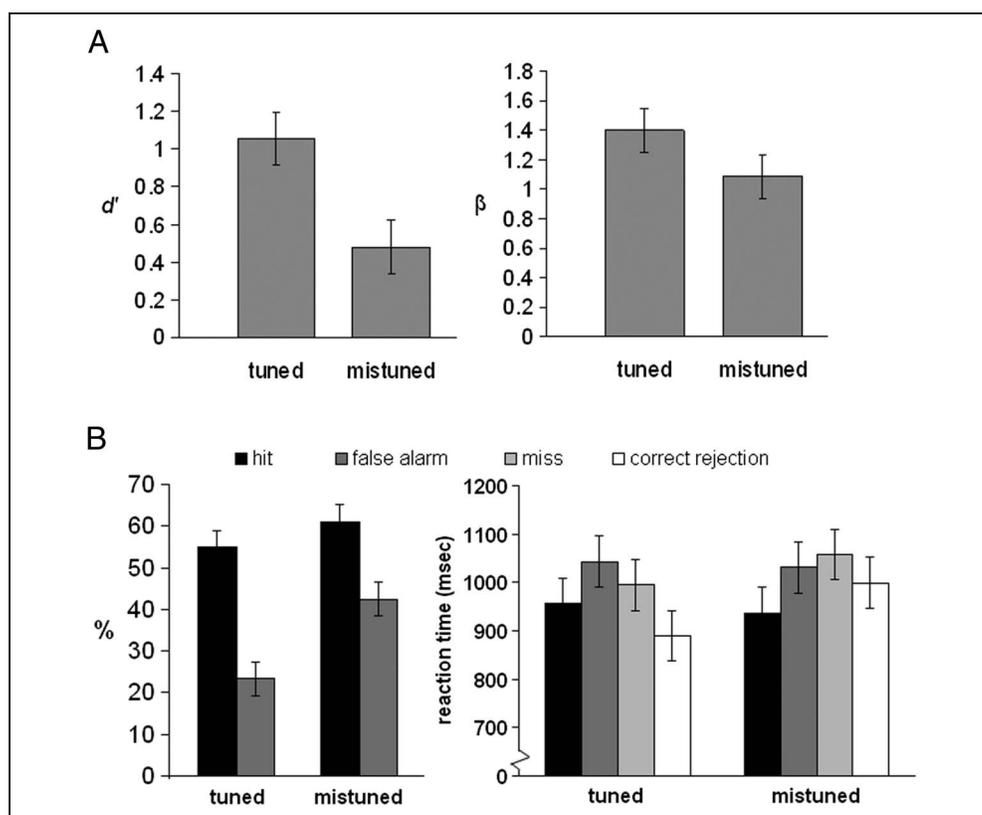
tuned), Gap (presence vs. absence), Response type (correct vs. incorrect), and Site (parieto-occipital electrode sites, namely, P1, Pz, P2, PO3, POz, PO4, O1, Oz, and O2). These nine electrodes best captured the LPC elicited by auditory stimuli (McDonald & Alain, 2005; Alain et al., 2001). Significant interactions between harmonicity, gap, and/or response type were examined using pairwise comparisons with Bonferroni adjustment. For all repeated-measures ANOVAs, we used the Greenhouse–Geisser to correct for any violations of assumption of sphericity (homogeneity of the covariance matrix).

## RESULTS

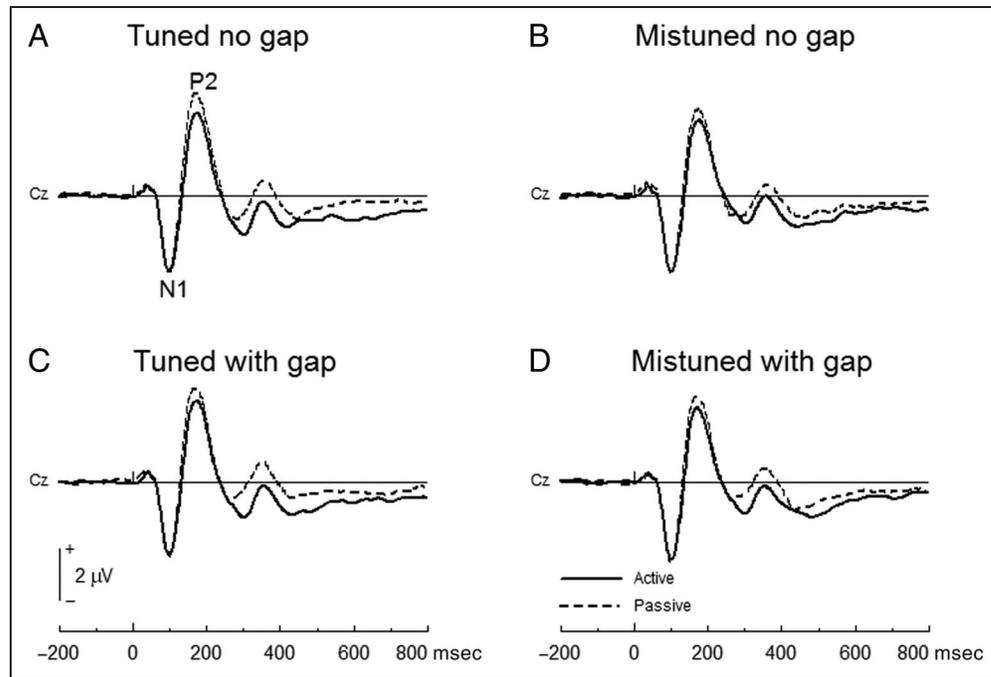
### Behavioral Response and Trial Types for ERP Analysis

Figure 1 shows the mean accuracy (mean  $d'$  and  $\beta$  values as well as hit and false alarm rates) and mean RT for the gap detection task. The hit rate was comparable across tuned and mistuned stimuli,  $t(19) = 1.01, p = .325$ , but the false alarm rate was higher for trials with mistuned stimuli,  $t(19) = 3.96, p < .001$ . Using the signal detection index, the mean  $d'$  was significantly lower for mistuned stimuli,  $t(19) = 3.98, p < .001$ . There was also a marginal effect of mistuning for the mean response bias,  $\beta$ ,  $t(19) = 2.01, p = .06$ . Participants were more prone to respond “gap” during trials with mistuned stimuli. For the tuned

**Figure 1.** Plot of  $d'$  and  $\beta$  (A) and hit rate and RT (B) for the gap detection task. Error bars indicate the SEM.



**Figure 2.** Plot of N1 and P2 waves of the active and passive listening conditions.



stimuli, the participants were more accurate in detecting gaps during trials with gap stimuli (Figure 1B).

RT was significantly shorter for correct than incorrect responses,  $F(1, 19) = 11.29, p = .0035$ , and shorter for tuned than mistuned stimuli,  $F(1, 19) = 4.81, p = .041$ . This pattern of results replicated the behavioral findings in Leung et al.'s (2011) study.

Overall, the number of trials available for ERP averaging was comparable among the eight trial types, with an average of  $136 \pm 45$  trials,  $F(1, 19) = 1.18-3.16, ps = .091-.291$ . However, there were slightly more trials where the participants correctly rejected the tuned stimuli ( $211 \pm 50$  trials) rather than the false alarms for tuned stimuli response ( $65 \pm 42$  trials).

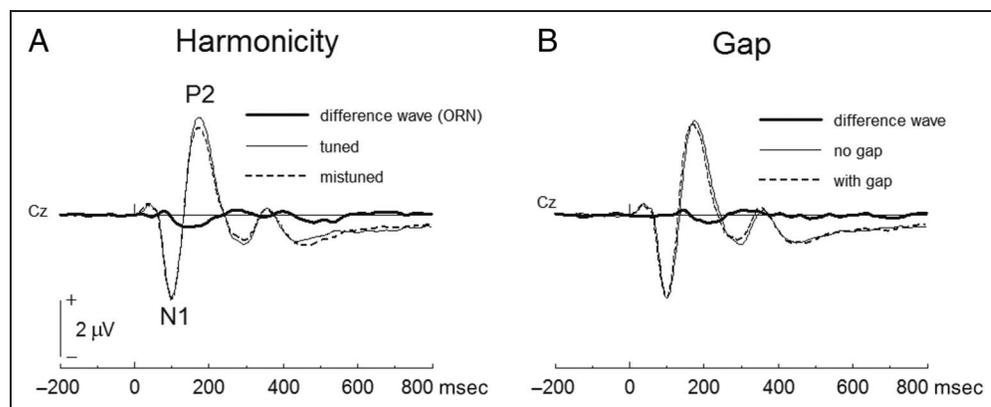
### Assessing the Encoding of Mistuning and Gap

Figure 2 shows the ERP waveforms for the active and passive listening conditions. For the mistuning compo-

nent (i.e., the ORN: mean amplitude between 100 and 200 msec), repeated-measures ANOVA (Condition  $\times$  Harmonicity  $\times$  Gap  $\times$  Site) revealed significant main effects of Condition,  $F(1, 19) = 8.79, p = .008$  (larger mean amplitude in passive than active conditions), and Harmonicity,  $F(1, 19) = 27.22, p < .001$  (larger mean amplitude in tuned than mistuned). Interaction effects among Condition, Harmonicity, and Gap were not significant,  $F(1, 19) = 0.16, p = .690$ , and interaction effects of these factors with Site were also not significant,  $F(8, 152) = 0.39-3.17, ps = .091-.927$ .

For the gap-related component (mean amplitude between 175 and 275 msec), repeated-measures ANOVA revealed significant main effects of Condition,  $F(1, 19) = 5.01, p = .037$  (larger mean amplitude in passive than active conditions), and Gap,  $F(1, 19) = 8.80, p = .008$  (larger mean amplitude in stimuli without a gap than stimuli with a gap). Interaction effects among Condition, Harmonicity, and Gap were not significant for both the mistuning and gap responses,  $F(1, 19) = 2.641,$

**Figure 3.** ERP waveforms at Cz for comparing between tuned and mistuned stimuli (A) and gap and no-gap stimuli (B). The waveforms are average of active and passive listening conditions. Difference waves show the ORN (A) and gap-related activity (B).

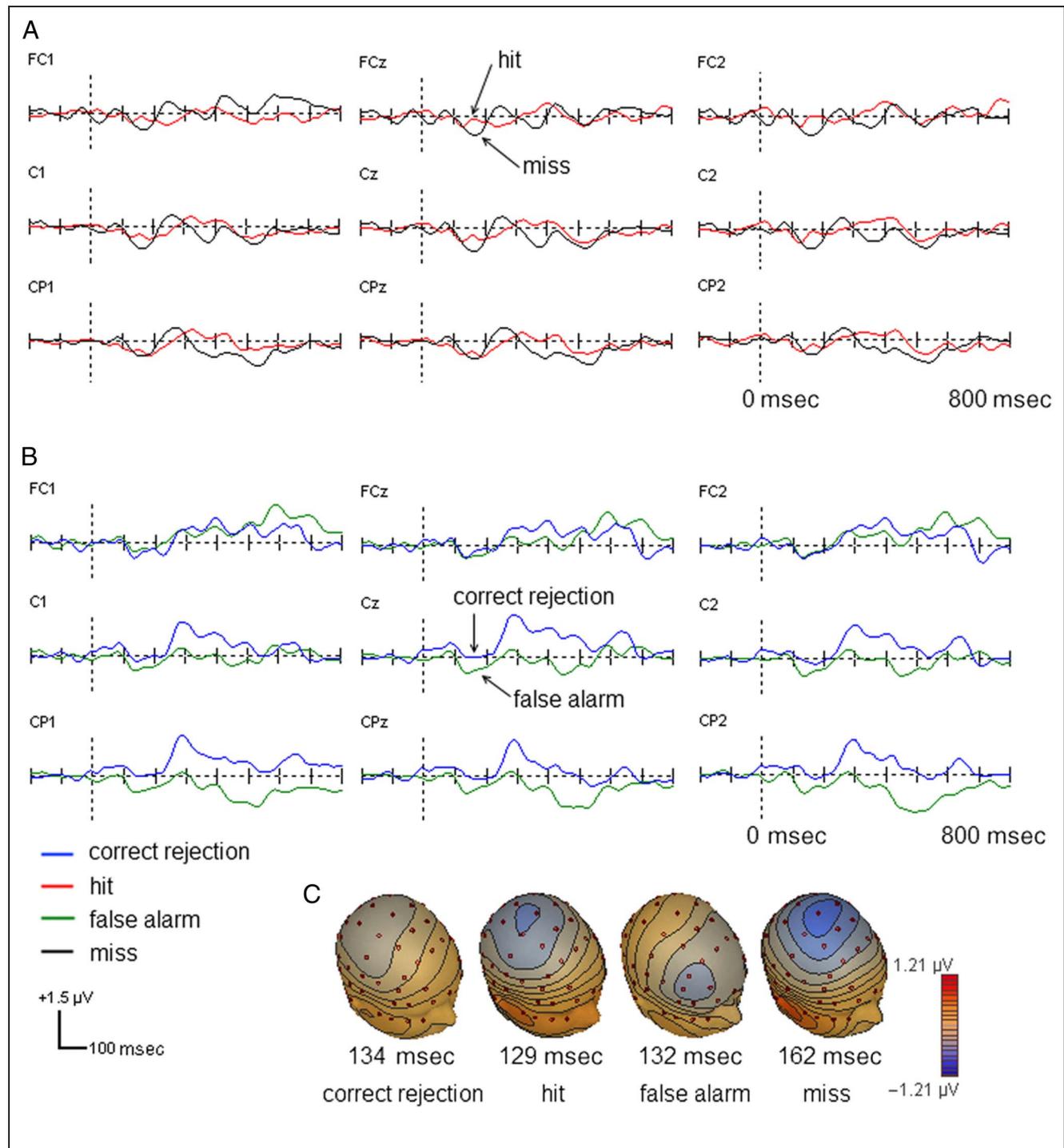


$p = .121$ , and interaction effects of these factors with Site were also not significant,  $F(8, 152) = 0.55-2.06$ ,  $ps = .168-.466$ . The results indicate that the mistuning and gap were independently encoded and the encoding process did not differ between the active and passive listening conditions. Figure 3 shows the ERP waveforms of the mistuning (the ORN component) and gap-related response.

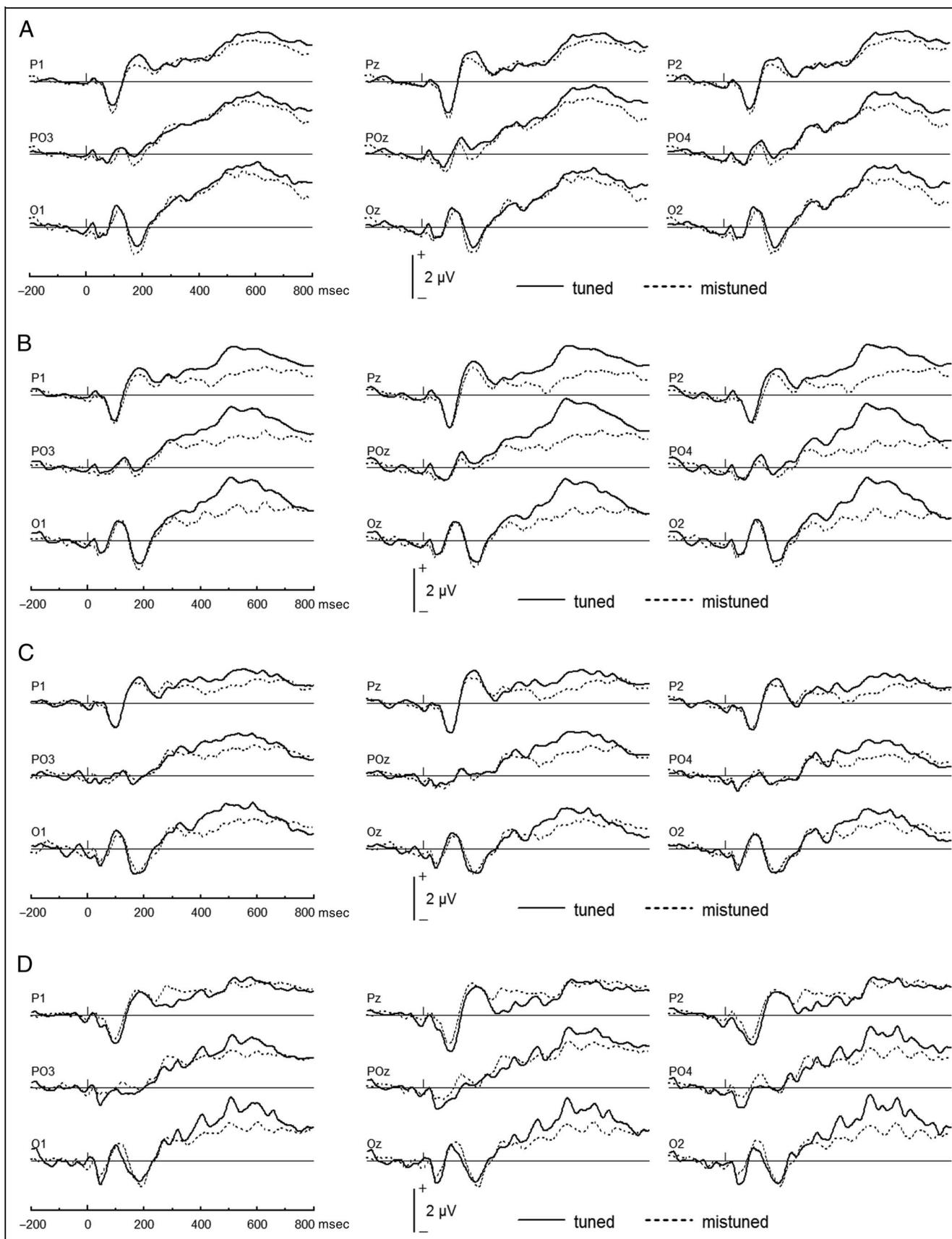
### Assessing Attention during Gap Detection

#### ANOVA Results for the ORN

Figure 4 shows the group mean ORN waveforms. Using the ORN (mean amplitude between 100 and 200 msec from the difference waves between the tuned and mistuned stimuli), repeated-measures ANOVA (Gap  $\times$  Response type) revealed a significant main effect of Response type,



**Figure 4.** Plots of the ORN for hit and miss responses (A) and correct rejection and false alarm (B) and their topography as a function of response (C).



**Figure 5.** Plots of LPC at parietal-occipital scalp locations, which compare tuned with mistuned stimuli for hit (A), correct rejection (B), miss (C), and false alarm (D).

$F(1, 19) = 11.60, p = .003$ , which indicated that the ORN was significantly smaller in correct than incorrect responses.

#### ANOVA Results for the Encoding of Gap during Gap Detection

To examine the interaction between successful gap detection and harmonicity, a repeated-measures ANOVA (Gap detection response [i.e., hit vs. miss]  $\times$  Harmonicity [i.e., tuned vs. mistuned]) on gap-related activity was performed. The interaction effect between Gap detection response and Harmonicity was not significant,  $F(1, 19) = .03, p = .871$ , which indicated that the mistuning did not cause any significant changes in low-level processing of the gap.

#### ANOVA Results for the LPC

The group mean LPC waveforms are shown in Figures 5 and 6. Overall, the LPC was larger for tuned than mistuned stimuli,  $F(1, 19) = 11.22, p < .0035$ , and for correct than incorrect responses,  $F(1, 19) = 9.22, p = .007$ . A three-way repeated-measures ANOVA (Harmonicity, Gap, and Response type) yielded significant interactions between Gap and Response type,  $F(1, 19) = 7.31, p = .015$ , and between Harmonicity, Gap, and Response type,  $F(1, 19) = 4.46, p = .049$ . Further pairwise comparisons were performed between experimental conditions.

To assess the impact of concurrent sound perception on attention, the LPC was compared between tuned and mistuned stimuli for different gap detection responses. For correct rejections, the LPC was greater for tuned than mistuned stimuli,  $F(1, 19) = 35.30, p < .0001$ , whereas for correct hits, the LPC was comparable for tuned and mistuned stimuli,  $F(1, 19) = 1.10, p = .307$  (Figure 5A and B). For misses, the LPC was greater for tuned than mistuned stimuli,  $F(1, 19) = 4.95, p = .0038$ , whereas for false alarms, the LPC amplitude was comparable for tuned and mistuned stimuli,  $F(1, 19) = .26, p = .616$  (Figure 5C and D).

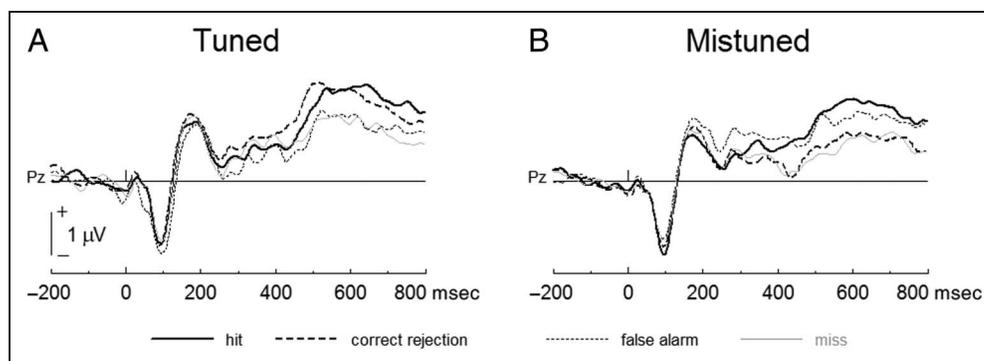
To address whether attention was related to successful gap detection, the LPC was compared between gap detection responses (i.e., hits vs. misses and correct re-

jections vs. false alarm) for tuned and mistuned stimuli. For tuned stimuli, the LPC was greater for hits than misses,  $F(1, 19) = 6.99, p = .016$ , and comparable for false alarms and correct rejections,  $F(1, 19) = 3.38, p = .082$  (Figure 6A). For mistuned stimuli, the LPC was greater for hits than misses,  $F(1, 19) = 8.39, p = .009$ , and greater for false alarms than correct rejections,  $F(1, 19) = 6.90, p = .017$  (Figure 6B). Additional significant results showed that the LPC was greater for correct rejections than misses for tuned stimuli,  $F(1, 19) = 31.20, p < .0001$ , for hits than correct rejections,  $F(1, 19) = 9.67, p = .006$ , and for false alarms than misses for mistuned stimuli,  $F(1, 19) = 14.75, p < .001$ .

## DISCUSSION

This study used auditory ERPs to determine whether gap detection impairments were because of the failure of low-level sensory encoding or the limitation of attentional capacity. The behavioral results of the current study are consistent with findings from a prior study by Leung et al. (2011) in which gap detection performance declined when mistuned harmonic stimuli were used. The present results delineate whether the failure to detect the gap is related to poor sensory encoding or taxed attentional processing. Here, the analyses of sensory evoked responses revealed comparable gap-related activity between tuned and mistuned harmonics during both active and passive listening conditions. In contrast, the analysis of attentional processes and response-related activity revealed main effects of both harmonicity and response accuracy as well as interaction effects among harmonicity, gap, and response accuracy. We found a marked decrease in the LPC amplitude when comparing incorrect with correct responses. In addition, the ORN, which indexes the perception of mistuned harmonics, was greater for incorrect than correct responses. The ORN has been proposed as a marker for the perception of concurrent sound objects, that is, of mistuned harmonics (e.g., Alain, 2007). A greater ORN for miss responses suggests that the perception of mistuned harmonic interfered with the perception of the gap in the sound complex. The results suggest that the presence

**Figure 6.** Comparing LPC between the trial types, namely, hit, correct rejection, miss, and false alarm, in each of the tuned (A) and mistuned (B) stimuli. Only Pz was illustrated for comparison as the waveform of each trial type was the same as Figure 5.



of a mistuned harmonic divided attention to more auditory objects, which in turn reduced successful gap detection.

In this study, the amplitude of the P2 wave was smaller for gap than no-gap stimuli. The reduced P2 amplitude could be accounted for by an N1 wave elicited by the onset of the gap that superimposed the P2 deflection generated by the earlier onset of the complex sound. A similar decrease in P2 amplitude has been reported in another study using similar gap durations (Heinrich et al., 2004). Importantly, this gap-related activity is distinct from the effects of mistuning as we found no interactions between the two until response-related evoked potentials. The fact that mistuning and gap effects do not interact statistically in the ERP components observed during the first 300 msec suggests that the lower detection rate for gaps presented in mistuned stimuli results from post-sensory-processing. We argue, therefore, that the reduced gap detection accuracy reflects limitations of attentional processing. Our findings could be explained by the object-based account of auditory scene analysis, which posits that attention operates on perceptual objects (Alain & Arnott, 2000).

Neisser (1967) proposed an object-based account of visual attention in which perceptual analysis takes place in two successive stages, a preattentive and a focal attention stage. The former is based on Gestalt properties to separate objects within a complex scene (Duncan, 1984). From this theory, the gap in this experiment was perceived as an entity having shared properties with the tuned harmonic sound, and the mistuned harmonic formed a separate auditory object that likely interfered with the perception of the tuned harmonic and gap. In the focal attention stage, attention is devoted to analyze a particular object in more detail (Neisser, 1967). However, previous studies have found that, even when a listener is sure which object is the target, object selection can fail when a competing object is inherently more salient, such as a much louder sound, than the target (Conway, Cowan, & Bunting, 2001). In our study, the most plausible explanation for this impairment of gap detection lies on the focal attention stage of the object-based theory. The mistuning appeared to be perceptually more salient than the gap because it occurs throughout the entire duration of the sound, whereas the gap only lasts a few milliseconds in the middle of the sound and is embedded in only 1 of the 10 tonal elements. Alain et al.'s (2001) study showed that, by mistuning a tone by 16%, like the one applied in this study, participants were very likely (over 95%) to perceive two sounds. However, the saliency of a gap could be affected by the frequency of the sound complex (Shailer & Moore, 1983, 1987). Some studies have found that the saliency of a gap in a tonal element could decrease because of the frequency difference between tonal elements of a complex sound (e.g., Oxenham, 2000). Therefore, the mistuned harmonic could compete more successfully for attention and prevent the gap from being detected.

One aspect of our ERP results supports this interpretation. Recall that we found a larger ORN on miss trials compared with hit trials. The suggestion is that, for misses, attention was more likely to be biased toward the more salient mistuned harmonic object, producing a larger ORN and drawing attention away from the gap. This argument is in line with the notion of biased competition, which proposes that the competition among representations happens before the volitional selection of objects as attention is involuntarily drawn through the salience of objects (Desimone & Duncan, 1995). In this study, it appears that the salience associated with the mistuned harmonic often wins the competition for representation presumably because the selective attention bias toward the gap is insufficient to override the interference from mistuning.

This study revealed differences in the LPC as a function of mistuning and response accuracy. Previous studies have shown that an increase in LPC amplitude typically reflects greater attentional resource allocation to a stimulus (Dien, Spencer, & Donchin, 2004; Solbakk, Reinvang, & Nielsen, 2000). Our data are consistent with this interpretation given that the LPC was larger during tuned than mistuned trials. For tuned stimuli, attention was initially allocated to the single tuned harmonic complex resulting in a large LPC waveform when the gap was correctly detected. In contrast, the presence of mistuning induces concurrent sound perception, which in turn increases the number of sound objects that compete for the limited attentional resource. Such a competition for attention among multiple auditory objects could account for the reduced LPC amplitude. This interpretation is consistent with prior studies showing a reduced LPC under the influence of competitive interference between concurrent stimuli (Martin, Jerger, & Mehta, 2007; Fischler & Bradley, 2006). This phenomenon is also consistent with Wickens's attentional resource model in which attention is divided among mental operations, which impairs performance task demands (Wickens, 1976, 2008).

The LPC is thought to reflect demands of attentional resources for the successful performance of the task (Starr & Don, 1988). Similarly, we found that the correctly responded trials elicited larger LPC than incorrectly responded trials for both the tuned and mistuned stimuli. This is consistent with previous studies that have reported positive relationships between behavioral response and LPC amplitude in auditory discrimination tasks (Xu et al., 2013).

Another consideration is that the gap is not only any stimulus but also the most critical stimulus for the task. Therefore, it is likely that attention was intentionally deployed so as to extract the gap of the sound in an optimal way. Prior studies have suggested that the LPC amplitude represents the amount of information extracted from the stimulus by the participants (Johnson, 1986). This might also be the case in this study as the LPC for hits was comparable between the tuned and mistuned stimuli, suggesting that attention might have been deployed to

extract the gap regardless of whether the distracting feature (i.e., the mistuning) was present.

One thing worth noting is that the gap was always embedded in the tuned component of the harmonic complex and never in the mistuned harmonic. The participants might have adopted the strategy to keep their attention focused on the tuned harmonic. Previous findings have shown that the LPC amplitude was greater in conditions where participants were asked to focus their attention to a particular stimulus than in conditions where focused attention to that particular stimulus was not required in active listening tasks (Martin et al., 2007). To verify this possibility, future studies may consider comparing LPC amplitudes during gap detection, with the gap having equal probability of being inserted in the tuned and mistuned components of a sound complex.

Last but not the least, it is noted that the ORN occurs at about the same latencies as the MMN component, which peaks between 150 and 250 msec after deviant onset (Näätänen, Paavilainen, Rinne, & Alho, 2007). The MMN has been used as an index of automatic change detection in the brain and is typically elicited by task-irrelevant sounds that occur infrequently in an otherwise regular stream of sounds. For example, it occurs when an incoming stimulus differs from the memory representation formed by the preceding stimulus sequence (Campbell, 2015; Näätänen et al., 2007). However, it is unlikely that the ORN was affected by the MMN in this experiment. One could imagine that a small MMN could occur after the occurrence of an inharmonic sound that would have followed a short series of repeated harmonic sounds. However, the opposite would be equally likely, given that the probability of presenting each type of sound was the same. As such, we would expect no net overall MMN effect for either type of stimuli.

## Conclusion

This study investigated whether the failure to detect gaps during concurrent sound perception is related to sensory encoding or attentional processing. During the first 300 msec, the ERP data showed comparable gap-related activity across the tuned and mistuned groups of harmonics. This suggests that the presence of a mistuned harmonic does not significantly alter early sensory responses of gap detection. Moreover, it suggests that the impaired gap detection is not likely because of the failure of sensory encoding of the gap. In contrast, the analysis of the late positive waves (LPC) revealed marked differences as a function of mistuning and response. These later effects suggest that the failure and success of detecting a gap are influenced by attentional resources. The results are consistent with the object-based account of auditory scene analysis in which a listener's attention is shared among multiple objects, which can lead to the reduction of attentional processing when compared with

situations in which attention is shared among fewer objects.

## Acknowledgments

This research was supported by grants from the Canadian Institutes of Health Research (MOP 106619) and the Natural Sciences and Engineering Research Council of Canada.

Reprint requests should be sent to Claude Alain, Rotman Research Institute, Baycrest Centre, 3560 Bathurst Street, Toronto, Ontario, Canada M6A 2E1, or via e-mail: calain@research.baycrest.org.

## REFERENCES

- Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*, *229*, 225–236.
- Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, *5*, D202–D212.
- Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1072–1089.
- Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, *15*, 1063–1073.
- Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *Journal of the Acoustical Society of America*, *111*, 990–995.
- Backer, K. C., & Alain, C. (2014). Attention to memory: Orienting attention to sound object representation. *Psychological Research*, *78*, 439–452.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Campbell, K., & Macdonald, M. (2011). The effects of attention and conscious state on the detection of gaps in long duration auditory stimuli. *Clinical Neurophysiology*, *122*, 738–747.
- Campbell, T. A. (2015). A theory of attentional modulations of the supratemporal generation of the auditory mismatch negativity (MMN). *Frontiers in Human Neuroscience*, *9*, e1065.
- Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin and Review*, *8*, 331–335.
- Curran, T., Schacter, D. L., Johnson, M. K., & Spinks, R. (2001). Brain potentials reflect behavioral differences in true and false recognition. *Journal of Cognitive Neuroscience*, *13*, 201–216.
- Cusack, R., Carlyon, R. P., & Robertson, I. H. (2000). Neglect between but not within auditory objects. *Journal of Cognitive Neuroscience*, *12*, 1056–1065.
- Dell'Acqua, R., Dux, P. E., Wyble, B., Doro, M., Sessa, P., Meconi, F., et al. (2015). The attentional blink impairs detection and delays encoding of visual information: Evidence from human electrophysiology. *Journal of Cognitive Neuroscience*, *27*, 720–735.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: Mental chronometry and the ERP

